included in all studies. Furthermore, the model with prior imputation of the variance appeared to be more stable than two-stage model.

PRM50

SAMPLE SIZE AND ETHICAL CONSIDERATIONS IN RANDOMIZED CLUSTER SAMPLING VERSUS INDIVIDUAL PATIENT RECRUITMENT FORMULAS IN PROSPECTIVE OBSERVATIONAL STUDIES

Chiappinelli R, Stanley S, Criscola R, Wasser T
HealthCore, Wilmington, DE, USA

OBJECTIVES: Differences in computational sample size formulas indicate that randomized cluster samples require more patients to demonstrate the same effect as studies that use Individual Patient Recruitment (IPR) formulas. We compared the differences in randomized cluster sampling and IPR formulas through a simulation study by varying the cluster size and Intra-Cluster Correlation Coefficient (ICC) to determine the magnitude of sample size differences. **METHODS:** The sample size formula for cluster sampling included two terms: 1) estimate of cluster size, and 2) estimate of ICC. Four Mean/Standard Deviation ratios were used reflecting the effect size, three ICC values, and three cluster sizes. Sample size was calculated for non-cluster and cluster formulas for 80% and 90% power. Sample size calculation results between cluster and IPR formulas were compared. **RESULTS:** Differences between cluster and IPR designs found that under sampling in IPR formulas vary from 5-15% and are largest when effect sizes are smallest. The IPR samples were smaller than cluster samples for the same effect size and power. Sample size using the cluster formula was smallest when ICC was small (0.15), at 80 percent power and cluster size of 5 patients per group. Cluster sample size was largest when ICC was large (0.25), at 90 percent power and cluster size of 20. **CONCLUSIONS:** In the research environment where prospective observational methods are used to gather "real world" data, studies that are conducted using cluster sampling, but powered with IPR formulas, are underpowered by as much as 15%. Ethical implications must be considered in prospective studies that require patient informed consent if the study is underpowered. If the prospective study involves risk the equipoise argument may be violated and place patients at risk (assuming there is a study treatment regimen), as the study may not be conclusive because of low power.

PRM51

NONPARAMETRIC REGRESSION ANALYSIS CONTROLS COST ANALYSIS IN DATA WITH OUTLIERS

Eisenberg D, Wasser T
HealthCore, Inc., Wilmington, DE, USA

OBJECTIVES: Cost analysis is often complicated to analyze because of skewed data caused by outliers in the upper tail of the distribution. Some of these outlier expenses are a result of extreme expenses before an observation period starts or during an episode of illness. Theil regression is a non-parametric linear regression method that provides accurate estimates of slope and intercept when outliers are present by calculating values based on the median. **METHODS:** In a study intended to measure the length of time it took for patient costs to return to normal pre-episode costs after pneumonia, the Theil method was used and compared to Ordinary Least Squares (OLS) results on the same data. The baseline cost was computed as the mean cost for the six months prior to diagnosis, the study allowed for a three month episode period and the OLS and Theil regression methods were computed on the monthly costs for the six months after the episode. **RESULTS:** High cost outliers during the three month episode led to elevated costs for the first post episode period. This caused an underestimate of cost using the OLS method. Theil regression correctly estimated the increased time to return to normal in 11 of the 21 variables tracked. These differences ranged from 15 to 370 days. OLS found extended time over Theil for 5 of 21 comparisons. These differences ranged from 2 to 26 days. Agreement between OLS and Theil was found for 5 of 21 comparisons. **CONCLUSIONS:** Outliers in regression analysis frequently occur when the variable of interest is cost. Theil regression offers considerable advantages over OLS regression when the outlier is in one of the tails of the distribution. The advantages include more accurate results as well being able to use all the data without exclusion of any data elements.

PRM52

SOCIOECONOMIC PATTERNS AMONG INTERNATIONAL IMMIGRANTS IN CHILE: THE USE OF CLUSTERS

Cabieses B, Tunstall H, Pickett K, Espinoza M
University of York, York, Yorkshire, England

OBJECTIVES: International immigration to Chile has increased in the past decade. Preliminary analysis found the immigrants were a very heterogeneous and polarized group in their SES which makes it difficult to identify particular needs of vulnerable subgroups within the total immigrant population. This analysis aims to describe their SES patterns. **METHODS:** Cross-sectional Chilean survey (CASEN-2006). From 268,873 participants, one percent were immigrants (n=1.877). Cluster analysis identifies subsets of a data set that contain similar points. Replacing these subsets by their aggregate properties, it creates a compact representation of the data set as a group of clusters. Hierarchical clustering is a step-wise process that merges the two closest or furthest data points or groups of data points at each step. Among the different types of hierarchical cluster analyses available, complete-linkage method was chosen as it creates clusters from the most distant values of the selected attributes (income, education and employment-status). Each SES-cluster was analysed in its demographic (age/sex/marital-status), geographical (urban-rural/region), SES variables (income/education/occupation), material-standards (overcrowding/sanitary-conditions/housing-quality). Analysis in STATA 10.0. **RESULTS:** After conducting complete-linkage hierarchical cluster analysis, three groups were identified: High-SES (n=398), Medium-SES (n=889), Low-SES (n=587). Key patterns are: High-SES: mean 35 years-old, 90% of working age, most married, technical or university level, only 2.7% with ethnic background. Medium-SES: mean 33 years-old, >60% technical education, mixed cluster. Low-SES: mean 25 years-old, >60% women, 8% ethnic background, up to high-school only, 2 poorest income quintiles. **CONCLUSIONS:** Immigrants in Chile are a very heterogeneous group, polarized by their SES. Hierarchical cluster analysis provided an appropriate method to group immigrants according to their socio-economic characteristics and, consequently, to provide clear patterns of SES vulnerability within the total immigrant population. Immigrants living in the Low-SES cluster are a vulnerable group that needs further attention in Chile.

PRM53

COMPARING MULTIPLE PROPENSITY SCORE ADJUSTMENT AND TRADITIONAL REGRESSION ANALYSIS TO ASSESS THE EXPOSURE-OUTCOME ASSOCIATION USING RETROSPECTIVE CLAIMS DATA

Chatterjee S, Aparasu RR, Chen H, Johnson M
University of Houston, Houston, TX, USA

OBJECTIVES: Researchers have suggested that, propensity score (PS) adjustment provides similar results as traditional regression analysis in observational studies. This has been attributed to the inappropriate implementation of PS, like inclusion of both PS and baseline covariates, and absence of covariate balance verification after PS adjustment. The present study employed a multiple PS adjustment model to evaluate the risk of falls/fractures in older adults using atypical antipsychotics, performed a balance check of covariates after PS adjustment and compared the results from multiple PS analysis with traditional regression model. **METHODS:** The study used IMS LifeLink Health Plan Claims Database and included older adults (aged ≥ 50 years) who initiated risperidone, olanzapine or quetiapine anytime during July 1, 2000 to June 30, 2008. Patients were followed until hospitalization/emergency room (ER) visit for falls/fractures, or end of the study period, whichever occurred earlier. Cox proportional hazard regression model was used to evaluate the relative risk of falls/fractures. The traditional model included over 80 baseline covariates which were also used to calculate the PS. The PS model included the two PS and their interaction terms. The covariate balance after PS adjustment was checked using logistic regression. **RESULTS:** After PS adjustment, there was no difference in any of the baseline covariates among the treatment groups. Both traditional regression and PS analyses had similar findings. There was no statistically significant difference with use of risperidone (Traditional: Hazard Ratio, HR, 1.10, 95% CI, 0.86-1.39; PS: HR, 1.09, 0.86-1.38) or quetiapine (Traditional: HR, 1.10, 0.84-1.44; PS: HR, 1.12, 0.86-1.46) compared to olanzapine in the risk of falls/fractures. **CONCLUSIONS:** The study findings suggest that, a PS adjustment model with well-balanced covariates across treatment groups gives similar results as traditional regression model.

PRM54

MODEL AND COVARIATE VISUALIZATION AIDS FOR ENHANCING THE INTERPRETATION OF STEPS IN THE HIGH DIMENSIONAL PROPENSITY SCORING ADJUSTMENT PROCEDURE

Juneau P[1], Huse DM[2]
[1]*Thomson Reuters Healthcare, Boyds, MD, USA,* [2]*Thomson Reuters, Cambridge, MA, USA*

OBJECTIVES: Currently, the work of Schneeweiss, et al. (2009) for propensity score adjustment is considered the standard approach for accounting for confounding in large claims data sets and is endorsed by such bodies as the Observational Medical Outcomes Partnership (OMOP) in the United States. The procedure appears to perform well and has many attractive features for the practitioner; however, examination of the selection of a set of potential effects for adjustment typically involves the perusal of large tables of summary statistics. For large data sets with potentially hundreds of covariates, this display does not afford the practitioner an easy, intuitive view of the relationships amongst the cofounders and with the desired outcome under study. **METHODS:** Modification of simple categorical data visualizations suggested by Cleveland (1993), Keller and Keller (1993), Harris (1999), Friendly (2001) and others were developed in common statistical software packages (e.g. SAS). **RESULTS:** The individual and joint behavior of the contribution of various confounders could be identified quickly and enhanced the user's understanding of their role in the procedure. **CONCLUSIONS:** In a setting with a large number of confounders, the procedure suggested by Schneeweiss, et al. reduces the number of confounders to a more manageable and practical level. Graphical techniques help the practitioner achieve a better understanding of the role of these confounders and the rationale for their inclusion in the adjustment procedure.

Research On Methods – Conceptual Papers

PRM55

A NEW APPROACH TO MODELING CANCER RECURRENCE AND FOLLOW-UP

Rose J[1], Augestad KM[2], Cooper G[1], Meropol NJ[1]
[1]*University Hospitals Case Medical Center, Cleveland, OH, USA,* [2]*University Hospital North Norway, Tromsoe, Norway*

OBJECTIVES: The ability to model cancer recurrence could assist in the optimization of surveillance strategies. However, capturing the dynamics of cancer recurrence in order to simulate follow-up surveillance after initial extirpative surgery presents a significant methodological challenge. The difficulty of modeling recurrence patterns is that relevant experimental and observational data is collected in the context of heterogeneous protocols for follow-up. Using the example of colorectal cancer, we propose a method of controlling for choice of follow-up regimen in order to infer the value of key natural history parameters. Once these values are inferred, any hypothetical follow-up regimen can be superimposed upon the nat-